Complexity vs. Performance: Empirical Analysis of Machine Learning as a Service

Yuanshun Yao, Zhujun Xiao, Bolun Wang*, Bimal Viswanath, Haitao Zheng and Ben Y. Zhao

The University of Chicago
*University of California, Santa Barbara

ysyao@cs.uchicago.edu
ML in Network Research

<table>
<thead>
<tr>
<th>congestion control protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Sivaraman et al., SIGCOMM’14</td>
</tr>
<tr>
<td>• Winstead &amp; Balakrishnan, SIGCOMM’13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>network link prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Liu et al., IMC’16</td>
</tr>
<tr>
<td>• Zhao et al., IMC’12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>user behavior analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Wang et al., IMC’14</td>
</tr>
<tr>
<td>• Zannettou et al., IMC’17</td>
</tr>
</tbody>
</table>

...
Running ML is Hard

Solution:
Machine Learning as a Service (ML-as-a-Service)
ML-as-a-Service

training data

user input
(model, parameter etc.)
Why Study ML-as-a-Service?

Q: How well do they perform?

Q: How much does the amount of user control impact ML performance?

Is my model good enough?
# ML-as-a-Service Platforms

<table>
<thead>
<tr>
<th>Google Prediction</th>
<th>Amazon ML</th>
<th>Microsoft ML</th>
<th>ABM</th>
<th>BigML</th>
<th>PIO</th>
</tr>
</thead>
</table>

- less amount of user input
- more
Control in ML
Control in ML

Data Cleaning

- Invalid/dup/missing data

training data

trained model
Control in ML

- **Data Cleaning**
  - Invalid/dup/missing data

- **Feature Selection**
  - Mutual Info, Pearson, Chi...

training data → trained model
Control in ML

Data Cleaning
- Invalid/dup/missing data

Classifier Choice
- Logistic Regression,
  Decision Tree, kNN...

Feature Selection
- Mutual Info, Pearson,
  Chi_square...

training data

trained model
Control in ML

Data Cleaning
- Invalid/dup/missing data

Classifier Choice
- Logistic Regression, Decision Tree, kNN...

Feature Selection
- Mutual Info, Pearson, Chi_square...

Parameter Tuning
- Logistic Regression: L1, L2, max_iter...

Training data → trained model
Control in ML-as-a-Service

Complexity vs. Performance?

Data Cleaning

Tuning

Google  ABM  Amazon  PIO  BigML  Microsoft

low  user control/complexity  high
Performance Measurement
Characterizing Performance

• Theoretical modeling is hard
  • Output of ML model depends on dataset
  • No access to implementation details

• Empirical data-driven analysis
  • Simulate a real-world scenario from end to end
  • Need a large number of diverse datasets

• Focus on binary classification
Dataset

• 119 datasets
  • From diverse application domains
  • Sample size: 15 - 245K, number of features: 1 - 4K
  • 79% of them are from UCI ML Repository

![Pie chart showing the distribution of datasets across various domains.]

- Life Science: 37%
- Computer Applications: 15%
- Artificial Test: 14%
- Social Science: 9%
- Physical Science: 8%
- Financial & Business: 6%
- Other: 11%
Methodology

• Tune all available control dimensions

Feature Selection

Classifier Choice

Parameter Tuning

training data

API

- Logistic Regression
- KNN
- SVM
- ...

API

- L1_reg
- L2_reg
- Max_iter
- ...

API

trained model
Methodology

• Tune all available control dimensions

Feature Selection  Classifier Choice  Parameter Tuning

training data  API  trained model  testing data  API
Trade-offs between Complexity and Performance
Complexity vs. Performance

Q: How does the complexity correlate with performance?
   - High complexity -> high performance
Complexity vs. Risk

• Q: How does the risk correlate with complexity?
  • High complexity -> high risk
Understanding Server-side Optimization
Reverse-engineering Optimization

• Q: Does server-side adapt to different datasets?

• Reverse-engineering using datasets
  • Create synthetic datasets
  • Use prediction results to infer classifier information
• Google switches between classifiers based on the dataset
• Use supervised learning to infer classifier family used
Takeaways

• ML-as-a-Service is an attractive tool to reduce workload

• But user control still has a large impact on performance

• Fully automated systems are less risky
Thank you!

Questions?